Irish Attitudes to Privacy in COVID-19 Times: Sentiment Analysis on Twitter and Survey Data

Contact tracing apps used in tracing and mitigating the spread of COVID-19 have sparked discussions and controversies worldwide. The major concerns in relation to these apps are around privacy. Ireland was in general praised for the design of its COVID tracker app, and the transparency through which privacy issues were addressed. However, the "voice" of the Irish public was not really heard or analysed. This study aimed to analyse the Irish public sentiment towards privacy and COVID tracker app. For this purpose we have conducted sentiment analysis on Twitter data collected from public Twitter accounts from Republic of Ireland. We collected COVID-19 related tweets generated in Ireland over a period of time from January 1, 2020 up to December 31, 2020 in order to perform sentiment analysis on this data set. Moreover, the study performed sentiment analysis on the feedback received from a national survey on privacy conducted in Republic of Ireland. The findings of the study reveal a significant criticism towards the app that relate to privacy concerns, but other aspects of the app as well. The findings also reveal some positive attitude towards the fight against COVID-19, but these are not necessarily related to the technological solutions employed for this purpose.

Additional Key Words and Phrases: privacy, Sentiment Analysis, COVID-19, tracking technology

ACM Reference Format:

. 2018. Irish Attitudes to Privacy in COVID-19 Times: Sentiment Analysis on Twitter and Survey Data. *Proc. ACM Meas. Anal. Comput. Syst.* 37, 4, Article 111 (August 2018), 11 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Contact tracing apps used in tracing and mitigating the spread of COVID-19 have sparked discussions and controversies worldwide. The major concerns in relation to these apps are around privacy. Ireland was praised for the design of its COVID tracker app, and the transparency through which NearForm[1] and the HSE (Ireland's Health Services) addressed privacy issues. The HSE provided a considerable amount of information on their website regarding the data processing, and made the data protection impact assessment (DPIA) of the app available to the public. The source code of the app is also available as open source and can be examined. However, concerns about privacy were raised for instance by the Irish Council for Civil Liberties due to lack of transparency from Apple and Google's side in terms of their involvement in the tracker app[2]. A research study also revealed issues with the DPIA and some of the documentation, some of the statements in the DPIA being shown as rather assumed than demonstrated[5]. The question is what is the Irish public sentiment? Are citizens concerned about

[1]https://www.nearform.com/
[2]https://www.iccl.ie/2020/experts-issue-pre-release-report-card-on-the-hse-covid-19-tracker-app/

Author's address:

2476-1249/2018/8-ART111 $15.00
https://doi.org/10.1145/1122445.1122456

their privacy ? Are these concerns preventing them from adopting the COVID tracker app? These are all very important questions as the success of the app is dependent on the percentage of people adopting it. Social media has grown drastically during the last few years. One of most widely used social media platforms is Twitter, where people from all over the world share short messages (within 280-character limit) on a wide range of topics. Starting from the beginning of 2020 till date, COVID-19 is among the most popular topics of discussion as it still continues to cause a worldwide pandemic. An unimaginable amount of COVID-19 related tweets has been and continues to be generated online. In this work, we perform an in-depth analysis of COVID-19 tweets generated in all over Ireland in order to investigate Irish people's attitude during this pandemic era. Our specific focus is on finding the attitude of the people living in Ireland towards tracking technology in general and COVID tracker app in particular and privacy. We also perform sentiment analysis on survey data, namely on participant's comments gathered in a survey on privacy during COVID-19 conducted in the Republic of Ireland[3]. The paper is organized as follows. Section 2 discusses the related work, section 3 focuses on the sentiment analysis on Twitter data, while section 4 present the analysis performed on the survey data. The final sections draws the conclusions.

2 RELATED WORK

Tracking technologies and apps used during COVID-19 have generated a lot of research. Much of this research perform technical analysis on the manner in which privacy is tackled[5]. Some other research focuses on conducting sentiment analysis in relation to these apps. Such a study was recently performed in relation to the HSE tracker app: a manual sentiment analysis on the Google Play comments [8]. The study focused mostly on the usability aspect. Our research goes beyond and performs sentiment analysis on data from two different sources: social media (Twitter) and survey data. We look broadly into the general attitude of the Irish people towards pandemic (pre- and during-) and then focus on tracker app and privacy.

A huge amount of research has been conducted in the area of social media analysis. One of the most explored area in this field is sentiment analysis that determines whether a social media user is talking about an event, a situation a product, service, or a brand in a positive, negative, or neutral way. Jayasanka et al. [4] analyse sentiments from Twitter social media by considering not only the polarity but also the use of those polarities in product profiling, trend analysis and forecasting. Samuels and Mcgonical [10] perform sentiment analysis of Twitter data on two subjects; (i) McDonalds, and (ii) KFC to show which restaurant has more popularity. Maheshwari et al. [6] investigate the application of Schwartz' psycholinguistic model of societal sentiment to social media data from Facebook and Twitter in order to identify the best linguistic and non-linguistic features for automatic classification of values and ethics. Balahur [2] presents a method for sentiment analysis specifically designed to work with tweets by considering their structure, length and specific language. Dos Santos and Gatti [3] apply a deep convolutional neural network that exploits from character- to sentence-level information in order to perform sentiment analysis of short texts. The work of Samuel et al. [9] demonstrate insights into the progress of fear-sentiment over time as COVID-19 approached peak levels in the United States. They provide a methodological overview of machine learning classification methods in the context of textual analytics and compare their effectiveness in classifying COVID-19 related Tweets of varying lengths. Our study is conceptually similar to the work of Mansoor et al. [7] that presents the global sentiment analysis of COVID-19 related tweets and how the sentiment of people in different countries has changed over time. However, we perform extensive analysis of the sentiment of Irish people during this pandemic with a focus on their sentiment towards tracing technology, tracker app and privacy. While our study may seem too localized, taking into consideration only Ireland, it is important to mention that Irish tracker app is the baseline for tracking apps in USA and other EU countries.

[3]https://www.rte.ie/brainstorm/2021/0203/1194741-covid-19-privacy-personal-data-tracker-app/

3 SENTIMENT ANALYSIS ON TWITTER DATA

We collected COVID-19 related tweets from Twitter public accounts of people living in Ireland. We crawled the COVID-19 tweets for the specific time period from 01/01/2020 (before the pandemic started in Ireland) up to 31/12/2020 (during ongoing pandemic). Our research methodology consists of four stages: (i) crawling of tweets, (ii) data cleaning, (iii) sentiment analysis .

3.1 Tweet Crawling

There are different methods to crawl tweets. The official way is to use Twitter API that needs a license from Twitter. Applying for a license needs cumbersome descriptions of your tasks and often takes a long time. Twint[4] is an open source project which can crawl tweets without time limits. It accepts keywords, language, geolocation, etc. as input parameters. For this reason, we decided to use this tool for crawling. Twint project accepts a configuration that can be adapted to a specific task. The most related configuration items are listed below:

- **Username**: Whose tweets you are going to crawl
- **Search**: Keywords that must be included in the tweets
- **Geo**: The Geolocation center and radius of region to be crawled
- **Lang**: the languages of the tweets
- **Output**: the file format
- **Since**: Start time of the tweets
- **Until**: End time of the tweets

The keywords defined for the tweet crawling first need to define some keywords for the tweets as following:

- coronavirus
- covid19
- covid19ireland
- covidtracker
- covidtrackerapp
- covidtrackerireland
- covid privacy
- covid tracker privacy
- covid tracker app privacy
- covid tracker app privacy attitude
- covid tracker app
- corona virus
- covid19 ireland

[4] https://github.com/twintproject/twint

We have chosen both general keywords (e.g. covid19) in the attempt to collect all relevant data, but also more specific to the problem investigated (e.g. covidtracker, covid privacy). There is a huge amount of Tweets about COVID-19, which comes from each corner of the world. In order to confine the search scope to Ireland, we chose several Geographic latitude and longitude coordinates and set approximate radius so that we can obtain the tweets mainly related to Ireland. We adopt four locations as the centre, using different radiuses to refine the crawling scope. The four locations are Dublin, Galway, Cork and county Donegal. Each location's Geographic coordinates and the radiuses are shown in Figure 1.

Fig. 1. Geolocations used as centre location to crawling tweets

We obtained more than two million tweets. Afterwards, we filter some noise and remove alignment mistakes, we obtain around 84 thousand tweets which is further reduced in number after removing duplicates and further cleaning that is described in the next section.

3.2 Data Cleaning

It was required to clean the data before performing the sentiment analysis. As tweets are user-generated content (UGC), they usually contain grammatical errors and violate linguistic norms. Due to the limit of 280 characters, in many cases, Twitter users tend to squeeze the contents of the messages by deliberately omitting unimportant words, shortening the sentences, words. This behaviour results in informal texts. In addition, many tweets contain hashtags, urls etc. that do not contribute much to the sentiment analysis process and make the process difficult. For example, consider the following tweet:

"Not So Fast - Ireland's Covid Tracker App https://t.co/E3gOLvokgk #COVIDTrackerApp #Location #DataPrivacy #DataProtection"

The above tweet contains an url (shown in blue) and few hashtags (shown in red). These entities usually create noise and impose challenges to the sentiment analysis process. We, therefore, clean the tweets by removing such contents. Hence, after cleaning, the above tweet will appear as follows: *"Not So Fast - Ireland's Covid Tracker App"*.

Table 1. Data statistics of Twitter data set

#Collected Tweets	#Tweets after cleaning	#Pre-pademic Tweets	#Pandemic Tweets
2, 495, 109	40, 820	1, 566	39, 254

Table 2. Performance comparison: TexBlob vs Afli

Tools used	Accuracy
TextBlob	64.81%
Afli	**65.74%**

3.3 Sentiment Analysis

In this section, we describe the sentiment analysis of the collected tweets. Twitter does not allow the crawling of all tweets and it limits the amount. However, we managed to crawl more than 2.49 million tweets with the help of "Twint" starting from 01/01/2020 until 31/12/2020. As the first lockdown in Ireland started on 12/03/2020, it is clear that this data contains tweets from both pre-pandemic and pandemic situations. We deliberately proceeded in this way in the attempt to capture a change in sentiment in the two periods of time, pre-pandemic and during pandemic. There were many identical tweets and many tweets with only hashtags, urls, etc. We removed these tweets to ease the process of sentiment analysis. As a result, only a part of tweets were considered for sentiment analysis. The data statistics is shown in Table 1. As can be seen from the table, the number of tweets has been drastically reduced after cleaning because of the identical tweets and a huge amount of noise. It can also be seen that most of the tweets are posted during the pandemic situation which is quite expected. There are 1, 566 pre-pandemic tweets in our data set. This is not unexpected because Irish people were not very much concerned until the pandemic started and so they did not post so many contents on Twitter. Although we have used many keywords to crawl the Twitter data, we found tweets for only following three keywords (i) covid19, (ii) covid19ireland, and (iii) covidprivacy. Note that there are no tweets relating to COVID tracker because the app did not exist in the pre-pandemic scenario.

3.3.1 Sentiment analysis tools. We used two different tools for sentiment analysis of the tweets: TextBlob and Afli's sentiment analysis tool.

Initially we used TextBlob[5] sentiment analysis tool. We tested it's performance on our test data set of 216 tweets that were manually annotated with three sentiment classes namely; (i) negative: showing dissatisfaction, unhappiness, compliant etc, (ii) neutral: generally consisting of queries, information regardless of sentiment, and (iii) positive: showing satisfaction, happiness, appreciation etc. When we tested the tool with 216 sentiment annotated tweets, it achieved an accuracy of 64.81% in classifying the sentiments correctly.

The sentiment analysis tool of Afli et al. [1] is specially designed for sentiment analysis in low resource scenarios, especially the tweets. This tool achieved over 70% of sentiment classification accuracy with their Twitter data set. When we tested the tool with our test data set of 216 sentiment annotated tweets, it achieved 65.74% of accuracy which is higher than that of TextBlob. The comparison is shown in Table 2. As Afli's tool exhibits better performance, we decided to use it for the sentiment classification of our main Twitter data set.

[5] https://pypi.org/project/textblob/

We performed sentiment analysis on the tweets from various angles. We looked at the pre-pandemic tweets, then we analysed the during pandemic tweets, then the overall tweets. A special attention was given to the tweets that were specifically related to the tracker application.

3.3.2 Analysis of pandemic related tweets. Figure 2 shows the partition of sentiment classes for the pre-pandemic, (in-)pandemic and all tweets together.

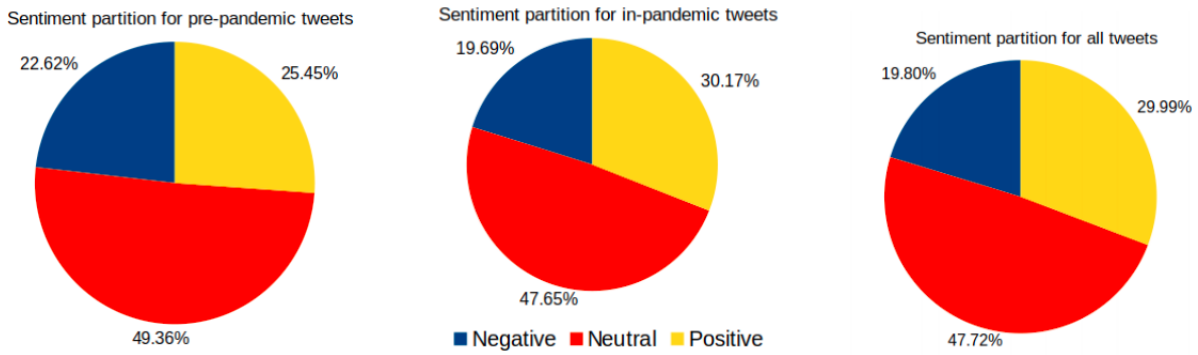

Fig. 2. Sentiment analysis of tweets in different scenarios

We notice in the figure that almost half (49.36%) of the tweets belong to the neutral category for the pre-pandemic tweets. This is less for the (in-)pandemic tweets which reflects similar change in all the tweets because the major part of the tweets belongs to pandemic times. However, the amount of negative tweets for the pandemic situation is less than that in the pre-pandemic situation (19% as compared to 22%). This means that people have posted more positive tweets than negative in the pandemic situation as compared to the pre-pandemic situation. We noticed that many people seem to be optimistic and say positive things about the fight against the virus. They appreciate the efforts of healthcare professionals, HSE(Ireland Health Services), Irish Government, healthcare professionals, and common people.

3.3.3 Analysis of tweets on COVID tracker application. There were not many tweets posted on the COVID tracker app in Irish social media, 1, 420 relevant tweets being collected. We have performed sentiment analysis on these tweets and the results can be seen in Figure 3. This figure shows the partition of sentiment classes of the tracker related tweets.

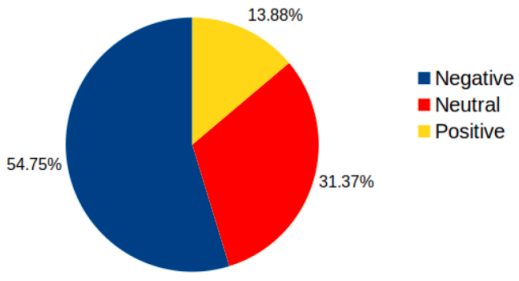

Fig. 3. Sentiment analysis of tweets on COVID tracker application

Table 3. Examples of correct and misclassification of pre-pandemic tweets

Example	Tweets	Predicted class	Expected class
1	"Again, superb coverage by @morningireland of #COVID19 Clear, calm, comprehensive. #publicservicebroadcasting #Coronavirusireland @rtenews"	Positive	Positive
2	@brendanoreillyp @NASCork Thanks for the kind support message. #PatientFirst #COVID19	Positive	Positive
3	that not make sense .The politicians canceled their meeting because they afraid of corona (but) when people ask them to cancel schools and universities their answer is ..... go to () #COVID19ireland	Negative	Negative
4	@simoncoveney @covid19 It's a pity you didn't stop thousands of Italians coming into Ireland last weekend. I expect to see a big leap in confirmed cases in two weeks.	Negative	Negative
5	Absolutely nothing on RTE about the patient who has been discharged after making a full recovery from #COVID19. Surprise surprised. #coronavirus	Positive	Negative
6	This while sad i think is clearly the right decision given the current threat that this virus poses. #FineGael #COVID2019 #marketcrash #COVID19ireland #coronvirusireland	Negative	Positive

It can be seen in the figure that only a 13.88% of the tweets are classified as positive with more than a half (54.7%) being classified as negative. Hence, our analysis shows that Irish people post more negative tweets than positive ones, which reflects their negative attitude towards the COVID tracker application. The tweets reveal privacy concerns, but other issues that relates to battery consumption, performance and efficacy.

3.4 Output Analysis

This section looks at the sentiment analysis output. Previously we discussed about the accuracy of the tools used and we motivated the decision of continuing our experiments with the tools that demonstrated highest accuracy. In looking at the output of the analsyis performed, we still noticed some misclassifications. We will give some examples in the next sections.

3.4.1 Output analysis of pre-pandemic tweets. Let us show some examples of both correct classification and misclassification of tweets in Table 3 below.

We can see in Table 3 that first 4 examples are correctly classified by the sentiment analysis system. In contrast, the last two examples (highlighted in red) are misclassified. It happens due to the complexity of the tweets. For example, the 5^{th} tweet in the table contains the negative segment (the segment with negative sentiment) "*Absolutely nothing on RTE ...*" and the positive segment (segment with positive sentiment) "*...making a full*

Table 4. Examples of correct and misclassification of pandemic tweets

Example	Tweets	Predicted class	Expected class
1	"Battery drain from Covid-19 app tracker https://t.co/fSY21Euqll	Negative	Negative
2	107 Clare people scheduled for COVID-19 swab failed to attend test over Christmas |	Negative	Negative
3	Wow they even have the #CovidTracker app translated to Irish already https://t.co/jXTfsJpiZH	Positive	Positive
4	"Covid Tracker app downloaded and running - for anyone concerned, you don't have to provide a phone number. You will be asked for it but it's optional. Check it out at https://t.co/sn6MSWrhA2 https://t.co/sk56kwm8yY"	Negative	Neutral
5	"Half a million deleted the Covid-19 tracker app after technical fault Problem is sorted out now though, so please download https://t.co/LSE6dK5atg Sent via @updayIE"	Negative	Neutral/Positive

recovery...". This tweet should actually be classified as a negative one because the main focus is on the first segment which is negative. However, the sentiment analysis tool gives more positive weight to the second segment than it gives negative weight to the first segment. It is, therefore misclassified as a positive tweet.

3.4.2 Output analysis of pandemic tweets. We now illustrate some examples of both correct classification and misclassification of pandemic tweets in Table 4 below.

First three rows in Table 4 show the examples of correct classification. On the contrary, the last two rows highlights some wrongly classified tweets. Consider example 5, this tweet is actually composed of two sentiment classes. The first part, i.e, " *Half a million deleted the Covid-19 tracker app after technical fault*" expresses negative opinion while the second part "*Problem is sorted out now though, so please download*" contains positive opinion. The tweet might be considered as neutral after combining these two sentiments. However, it may also belong to positive class because of the indication that the problem of the tracker app is solved eventually, as sometimes the overall sentiment class is decided based upon the underlying sentiment towards the end of the tweet. Therefore, this tweet should have been classified as either neutral or positive but it is misclassified as negative by the sentiment analysis system.

4 SENTIMENT ANALYSIS ON SURVEY DATA

A national survey on privacy was conducted in Republic of Ireland that was launched in November 2020 and was closed in the beginning of January 2021. The survey shown that Irish people while aware of their privacy

are more willing to share their data during the pandemic in the interest of saving lives[6]. The survey invited participants to leave feedback; out of 1012 participants, 202 left comments. Most of these comments are about the COVID tracker application itself. 30.7% of them are related to the survey and 69.3% are about the tracker app. To better understand the comments, we manually inspected all the comments and divided them into three classes depending upon the degree of underlying sentiment:

(i) negative: if the comment expresses dissatisfaction, unhappiness, complaint etc. (ii) neutral: if the comment is a query, information etc. regardless of any sentiment, (iii) positive: if the comment expresses satisfaction, happiness etc.

The feedback on the COVID tracker app are mostly negative: more than half (50.71% as it can be seen in Fig. 4) of the comments express negative opinions. In contrast, only 13.57% of the comments are positive about the application and imply that some people consider the application to be efficient in controlling the pandemic. The rest 35.71% of the comments are neutral, i.e, either they are simply providing information or are queries, regardless of any sentiment. These results on the survey data are extremely similar to the ones that were obtained for the tweets. Many negative comments about the tracker app are around privacy, for example: *I understand the benefits of information that privacy violating trackers can have in helping with the Covid crisis. However, it also carries a risk of enabling more unethical privacy violating data collection activities to be more "tolerated" and "seen as acceptable" after the Covid crisis has ceased.* This comment is labeling all tracker apps as apps that are clearly violating the privacy. In general, the negative comments seem to indicate that people perceive the tracker app as a gateway to surveillance of the population beyond COVID-19. Another comment states for instance: *The fight against COVID19 is important, but the line between fighting COVID19 and privacy is very thin.Could easily be exploited and profited off beyond the cause of fighting COVID19.* It is perceived as a dangerous precedent.

There are a few positive comments that relate to privacy and appreciate HSE's effort to be as transparent as possible (by making the data protection impact assessment available for instance) and the design of the app with respect to privacy. For instance one participant stated: emphI only use the app because I read the DPIA and was reassured by it, while another one stated: The COVID 19 tracker has been well designed with respect to privacy considerations

Another important issue that is highlighted in many comments is the issue of efficacy and the lack of communication in terms of how effective the app is. Some relevant comments in this regard are for instance:

emphI use the covid tracker but note that I hear little about how useful it is thought to have been - Silence after the initial hype. and emphUse of tracker system to date in Ireland UK(?): Real impact / success = ?

Other issues were signaled as well: battery issue, user experience-related and issues related to backward compatibility. The issue of battery consumption was a well-known issue of the HSE tracker app that was fixed in the meantime.

The general sentiment among the people using the app seems to be that they perceive the use of the app as a sacrifice of their privacy and they expect to see that their sacrifice counts. They want to see conclusive data about the efficacy of the app.

5 CONCLUSIONS

In this paper, we have performed sentiment analysis on he Irish tweets related to COVID-19 over quite a long period of time, before the pandemic started to manifest itself in Ireland, beginning of 2020 until the end of 2020, namely a full year. We have also performed sentiment analysis on survey data represented by the comments

[6]https://www.irishtimes.com/business/technology/covid-19-big-rise-in-inclination-to-share-data-due-to-pandemic-1.4478392

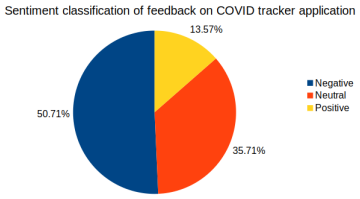

Fig. 4. Sentiment analysis of survey data on COVID tracker application

gathered in a survey on privacy during COVID-19 conducted in the Republic of Ireland. We had a specific focus on analysing the Irish public sentiment towards tracking technology in general and HSE COVID tracker app used in the Republic of Ireland in particular and privacy. Our results show that many people seems to be optimistic and say positive things about the fight against the virus. They also appreciate the efforts of HSE, healthcare professionals, and common people. They are not that positive in relation to the tracking technology and COVID tracker app. Both analysis, on tweets and survey data, unveils more a predominantly negative sentiment towards the tracker app. Privacy concerns are one of the main causes for this sentiment, but other issues are revealed as well, with efficacy being one of the most prominent among these. People appreciate the transparent communication (e.g. the publishing of the DPIA on the HSE impact, open discussion about privacy concerns), however the general sentiment even among the people using the app is that they are sacrificing their privacy. Hence, the importance of the other theme, the efficacy theme. As they feel that they are sacrificing their privacy, people need to see that their sacrifice matter, they want to see conclusive data about its efficacy.

ACKNOWLEDGMENTS

*****

REFERENCES

[1] Haithem Afli, Sorcha McGuire, and Andy Way. 2017. Sentiment Translation for low resourced languages: Experiments on Irish General Election Tweets. In *18th International Conference on Computational Linguistics and Intelligent Text Processing*. Budapest, Hungary.

[2] Alexandra Balahur. 2013. Sentiment Analysis in Social Media Texts. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Association for Computational Linguistics, Atlanta, Georgia, 120–128. https://www.aclweb.org/anthology/W13-1617

[3] Cícero Dos Santos and Maíra Gatti. 2014. Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. Dublin City University and Association for Computational Linguistics, Dublin, Ireland, 69–78. https://www.aclweb.org/anthology/C14-1008

[4] Sachira Jayasanka, Thilina Madhushani, E. Marcus, I. Aberathne, and Saminda Premaratne. 2013. Sentiment Analysis for Social Media. 1–11.

[5] Douglas J Leith and Stephen Farrell. 2020. Contact tracing app privacy: What data is shared by Europe's GAEN contact tracing apps. *Testing Apps for COVID-19 Tracing (TACT)* (2020).

[6] Tushar Maheshwari, Aishwarya N. Reganti, Samiksha Gupta, Anupam Jamatia, Upendra Kumar, Björn Gambäck, and Amitava Das. 2017. A Societal Sentiment Analysis: Predicting the Values and Ethics of Individuals by Analysing Social Media Content. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. Association for Computational Linguistics, Valencia, Spain, 731–741. https://www.aclweb.org/anthology/E17-1069

[7] Muvazima Mansoor, Kirthika Gurumurthy, Anantharam R U, and V R Badri Prasad. 2020. Global Sentiment Analysis Of COVID-19 Tweets Over Time. arXiv:2010.14234 [cs.CL]

[8] Kaavya Rekanar, Ian R O'Keeffe, Sarah Buckley, Manzar Abbas, Sarah Beecham, Muslim Chochlov, Brian Fitzgerald, Liam Glynn, Kevin Johnson, John Laffey, et al. 2021. Sentiment analysis of user feedback on the HSE's Covid-19 contact tracing app. *Irish Journal of Medical Science (1971-)* (2021), 1–10.

[9] Jim Samuel, G. G. Md. Nawaz Ali, Md. Mokhlesur Rahman, Ek Esawi, and Yana Samuel. 2020. COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification. *Information* 11, 6 (2020). https://www.mdpi.com/2078-2489/11/6/314

[10] Antony Samuels and John Mcgonical. 2020. Sentiment Analysis on Social Media Content. arXiv:2007.02144 [cs.CL]